

Design and implementation of a distributed data acquisition function architecture based on DOA/Handle technology

Hongfei Li¹, Yong Kong^{2*}, and Xueli Wang³

¹Department of Security, China Control Systems Cyber Emergency Response Team, Beijing, China

²Shenzhen Olym Information Security Technology Co., Ltd, Shenzhen, China

³School of Physics and Intelligent Manufacturing Engineering, Chifeng University, Chifeng, China

Abstract. With the development of new generation Internet technology, Digital Object Architecture (DOA)/Handle system plays an important role in industrial system. The development and technical characteristics of DOA/Handle technology is studied in this paper. Then function architecture of distributed data acquisition system based on DOA/Handle technology is proposed. Through the detailed design of the functions and characteristics of the architecture, the core functions such as distributed data acquisition, distributed data management, data security sharing, and data security governance are realized in the industrial Internet environment. At the end of the paper, the typical scenarios of this architecture are summarized, and a prospect of the application prospect of DOA/Handle technology is introduced.

1 Introduction

The Digital Object Architecture (DOA) was proposed by Dr. Robert Kahn, who is the founder of the Internet, Turing Award winner, and co-inventor of the TCP/IP protocol. The concept of DOA is to provide a set of Internet infrastructure to realize the connection, sharing and security management of information between various information systems. Handle system is the core realization of the DOA, which fully implements the concept and functional model of the DOA. Handle system realizes the comprehensive management capabilities of the four dimensions of digital object (DO) identification, analysis, information management, and information security, with powerful data management functions and complete security mechanisms. As an important technology of the new generation of Internet, it can not only support the construction of application platforms, but also realize safe interoperability between different places, different hosts, and heterogeneous information systems, which can be effective solve the problem of information islands and data chimneys on the Internet.

The function architecture of distributed data acquisition system based on DOA/Handle technology is proposed, referred to as Handle Distributed Data Acquisition Function Architecture (HDAFA) in this paper. HDAFA is extremely suitable for large-scale discrete

* Corresponding author: kongyong302@163.com

data collection and data object management scenarios. The functional implementation detail of HDAFA in the industrial Internet scenario are introduced. At the end of the paper, the typical scenarios of this architecture are summarized, and a prospect of the application prospect of DOA/Handle technology is introduced.

2 HDAFA design

2.1 Design goals

Data acquisition is a core part of large -sized application systems or application platform systems. Its meaning is to directly face all kinds of data objects for data collection and standardized sorting, and to provide a series of standardized data services for others in the overall system. The module is used for invocation, statistics and analysis, and plays a role in linking up the upstream and downstream information sharing and exchange.

The HDAFA architecture in this paper is designed and implemented using Handle technology, and various types of information discretely distributed on the network are standardized and defined, filtered and classified, sorted and summarized, and data management is achieved through a unified collection method; at the same time, support The core function of information security interoperability between remote, different host, and heterogeneous information systems in the industrial Internet environment.

2.2 Implementation logic

The HDAFA adopts a two-layer Handle node group model to realize the construction of logical structure and can provide virtual data resource pool services. as shown in Fig. 1.

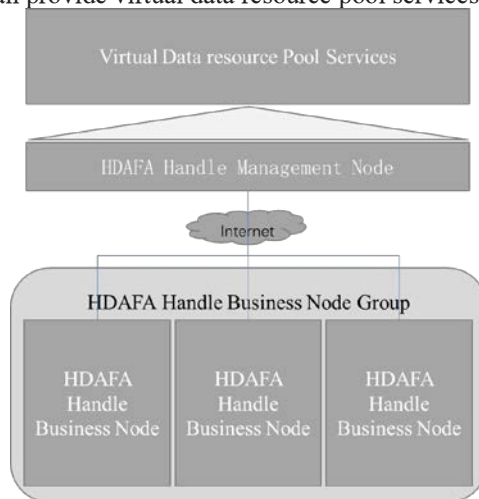


Fig. 1. Realization of the logical hierarchy structure diagram.

First of all, in the industrial Internet environment, the main design is to realize data collection through four main forms: the first is to connect the Handle system with major industrial Internet platforms in the industrial Internet field to connect the relevant industrial Internet platforms. Data is standardized data registration and localized information collection through metadata definition, classification, identification, etc.; the second is to dock various industrial enterprises that have not yet joined the industrial Internet platform, and dock the Handle system with its industry-related systems. Data definition, classification,

identification and other methods are used for standardized data registration and localized information collection; the third type is the industrial information data actively reported by companies with a low degree of informatization. After standardization and sorting, it is directly registered in the Handle system, To form a centralized collection; the fourth is to use web crawler technology to actively crawl the enterprise's industry-related public data, and register it in the Handle system after data sorting and processing.

Secondly, gather multiple docked "data collection Handle business nodes" and gradually form a "data collection Handle business node group", and build the upper "data collection Handle management node" to form a two-tier business management model of many-to-one, "Data Collection Handle Management Node" will uniformly manage and down-monitor the docked "Data Collection Handle Business Node Group", and perform various data management services through a unified business management model.

Finally, on the basis of the first two items, a virtualized data resource pool service is constructed as a whole, and externally embodied by providing diversified data interface services based on the standard Handle protocol to support external system integration and data utilization.

2.3 Functional architecture

2.3.1 Overall overview

In the specific architecture design, the bottom-up structure can be divided into three layers: data source layer, data collection layer, and data service layer. The overall functional architecture is shown in Fig. 2.

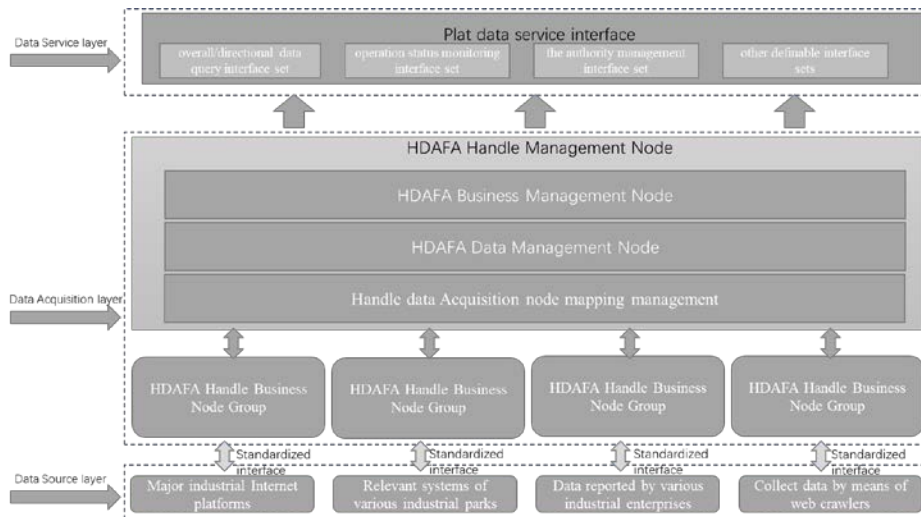


Fig. 2. Overall functional architecture diagram of the HDFAFA.

The data source layer mainly defines the data sources that need to be collected and sorted; the data acquisition layer mainly performs data collection related work including data classification definition, data collection, data sorting and other functions by connecting to the data source layer. At the same time, this layer also provide services such as unified business management and inter-level node management; the data service layer provides customized data interface services based on the data reserve provided by the data collection layer according to the needs of different modes.

2.3.2 Function realization in Layers

2.3.2.1 Data source layer

There are Major industrial Internet platforms, Relevant systems of various industrial parks, Data reported by various industrial enterprises, and Collect data by means of web crawlers in the Data source layer. Major industrial Internet platforms can provide the system with relatively standardized data sources of various enterprises under it. Relevant systems of various industrial parks can provide relatively standardized data sources for each enterprise in the park. Data reported by various industrial enterprises can Provide a unified registration entry management page through the Handle system, and standardize and enter data independently reported by industrial enterprises, including enterprise information, product information, and manufacturer information of various industrial enterprises. Collect data by means of web crawlers mainly include industry-related news, policies, public opinion and other data that are disclosed on the Internet, and the crawled data is further cleaned and processed, and the information is collected through the unified interface protocol of the Handle system. System data is registered to the Handle system.

2.3.2.2 Data acquisition layer

The Data acquisition layer is mainly designed and implemented in a two-layer mode of HDAFA Handle management node-HDAFA Handle business node group. Set up a Data acquisition Handle management node, unified management of data acquisition Handle node mapping information, and provide business management services, data acquisition and management services and other functions, as the main node of data acquisition and aggregation. Through the data monitoring of the HDAFA Handle management node, data monitoring and summary of downstream business nodes can be realized, and it can provide first-hand data support for other systems.

2.3.2.3 Data service layer

The Data service layer includes the overall/directional data query interface set, the operation status monitoring interface set, the authority management interface set and other definable interface sets.

3 HDAFA implementation and Applicable scenario analysis

A typical implementation of the HDAF architecture is to integrate with the industrial Internet data risk monitoring platform to provide relevant data and functional information for the platform. The goal is to build a cross-industry and cross-domain industrial Internet platform, public service support platform, and industrial Internet demonstration zone. Monitor and analyze the development trend of other countries. Service deployment data acquisition has more than 10 Handle business nodes, 1 management node, 1 backup node, and 1 management terminal. Standardized docking is carried out for 4 different data sources, and the functional architecture in this article is fully realized and achieved good results. effect.

As a general-purpose basic functional architecture, the HDAFA has a wide range of applicable scenarios. It is suitable for the acquisition of various types of discrete data, such as: group enterprises distributed in various places, without spending a lot of resources on centralized data centers. Construction and management, using this data acquisition

functional architecture, on the basis of not affecting the original enterprise group information architecture, can quickly build an effective data acquisition management platform, and can be compared with other companies in the original enterprise. The system integrates organically and on-demand. It is worth mentioning that this combination and integration is a lightweight and standardized integration that can be integrated without affecting the normal operation of the original system business.

4 Conclusion

The distributed data acquisition function architecture based on DOA/Handle technology (HDAFA) focuses on the acquisition of discrete information and data on the network. Using the distributed characteristics of the Handle system to achieve a two-layer structure of distributed acquisition, centralized management, the collected data is classified and stored: centralized storage of key data information, distributed local storage of other data, greatly reducing traditional methods. The large number of data streams generated during the following data acquisition process greatly saves network bandwidth resources, while greatly reducing the magnitude and difficulty of later centralized data management. Using Handle system data object metadata standard reference and custom mechanism to ensure data. The owner has independent data management authority, and can also conveniently complete the definition, cleaning and sorting process of the data format in accordance with the unified metadata standard. Using Handle system data confirmation and safety management mechanism to realize the safe transmission of data during the collection of data. The use of Handle system security mechanism ensures the safe management and interaction of data, and at the same time dispels the worries of enterprises in the process of data exchange and sharing.

I would like to express my gratitude to all those who helped me during the writing of this paper. Grateful acknowledgement is made to Dr. Yong Kong, my dear friend as well as my teacher. Finally, this work was financially supported by National Key Research and Development Program of China (2018YFB2100400).

References

1. P Mockapetris. Domain Names - implementation and specification. RFC1035, November, 1987, 4.
2. P Faltstrom, E.164 number and DNS. RFC2916, september, 2000, 1-4.
3. Sam X Sun. Internationalization of the Handle system - A Persistent Global Name service. Proceeding of 12th Inter- national Unicode Conference, April, 1998.
4. Kahn, Robert and Wilensky, Robert. A Framework for Distributed Digital Object services. May, 1995.
5. IETF RFC 3650. Handle System Overview[S].
6. IETF RFC 3651.Handle System Namespace and Service Definition[S].
7. KAHN R. The Digital Object Architecture and its application to the Internet of things. Beijing, 2010.
8. GUO X, LI Y SUN S X. Federated Content Rights Management for Research and Academic Publications Using the Handle System [J]. D-LibMagazine, 2010,16 (11/12).